Standard Error of Volume Forecast using Multiple Linear Regression Model

Wuben Luo, Ph.D, P.Eng.

BC Hydro Resource Management

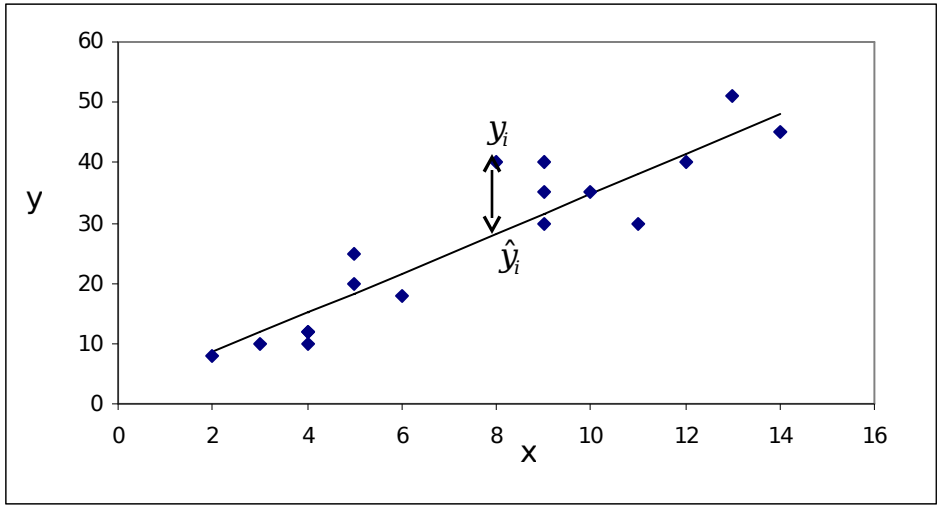
Agenda

- Volume Forecast based on MLR
- Standard Error of Residuals (s_{Resid})
- Standard Error of Prediction (s_{Pred})
- Problem with Current Practice of Using s_{Resid} for s_{Pred}
- Calculation of s_{Pred}
- Confidence Interval
- An Example in Excel
- Summary

Runoff Volume Forecast Based on MLR Model

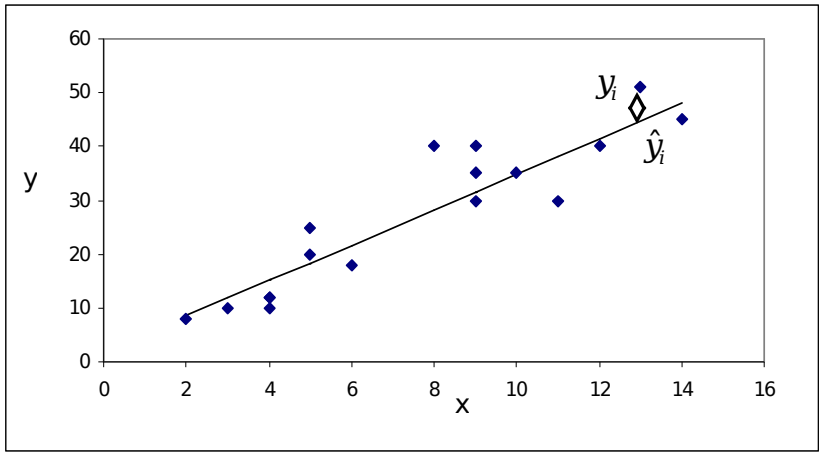
- **MLR:** a statistical procedure based on minimizing summed squared errors
- **Dependent Variable:** residual runoff volume
  - For February (or forecast date) to September
  - Snowmelt dominated
- **Independent Variables:**
  - Snow water equivalent
  - Accumulated winter precipitation
  - Antecedent temperature
- **Historical data availability:**
  - ~ 30 years

Standard Error of Residuals (S_{Resid})



- It is a random error
- Represents inherent variation in volume forecast
- Constant for the whole range of *y*

Standard Error of Prediction (s_{Pred})



- **Modeling error:**
  - due to use of a sample of limited size
  - Minimal when *x* is at the mean; Increases as *x* deviates from mean
  - can be reduced with better information
- **s_{Pred}:** is a combination of residual error and modeling error

Current practice of using s_{Resid} for s_{Pred}

- Accounts for only the random error
- Constant for all *y*
- Underestimates the total error of prediction
  - Underestimation is minimal at the mean of *x's*
  - Underestimation becomes significant as *x's* deviate from means
- Overestimates the confidence interval
- Has adverse effect on reservoir operations planning particularly in extreme weather years.

Calculation of s_{Pred} - Simple Regression Model

$$s_{Pred_y|x_0} = s_{Resid}\sqrt{1+\frac{1}{n}\left[1+\frac{(x_0 - \bar{x})^2}{s_x^2}\right]}$$

Calculation of s_{Pred} - MLR

$$S_{Pred_y|x_0} = S_{Resid}\sqrt{(1+\mathbf{x_0}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x_0})}$$

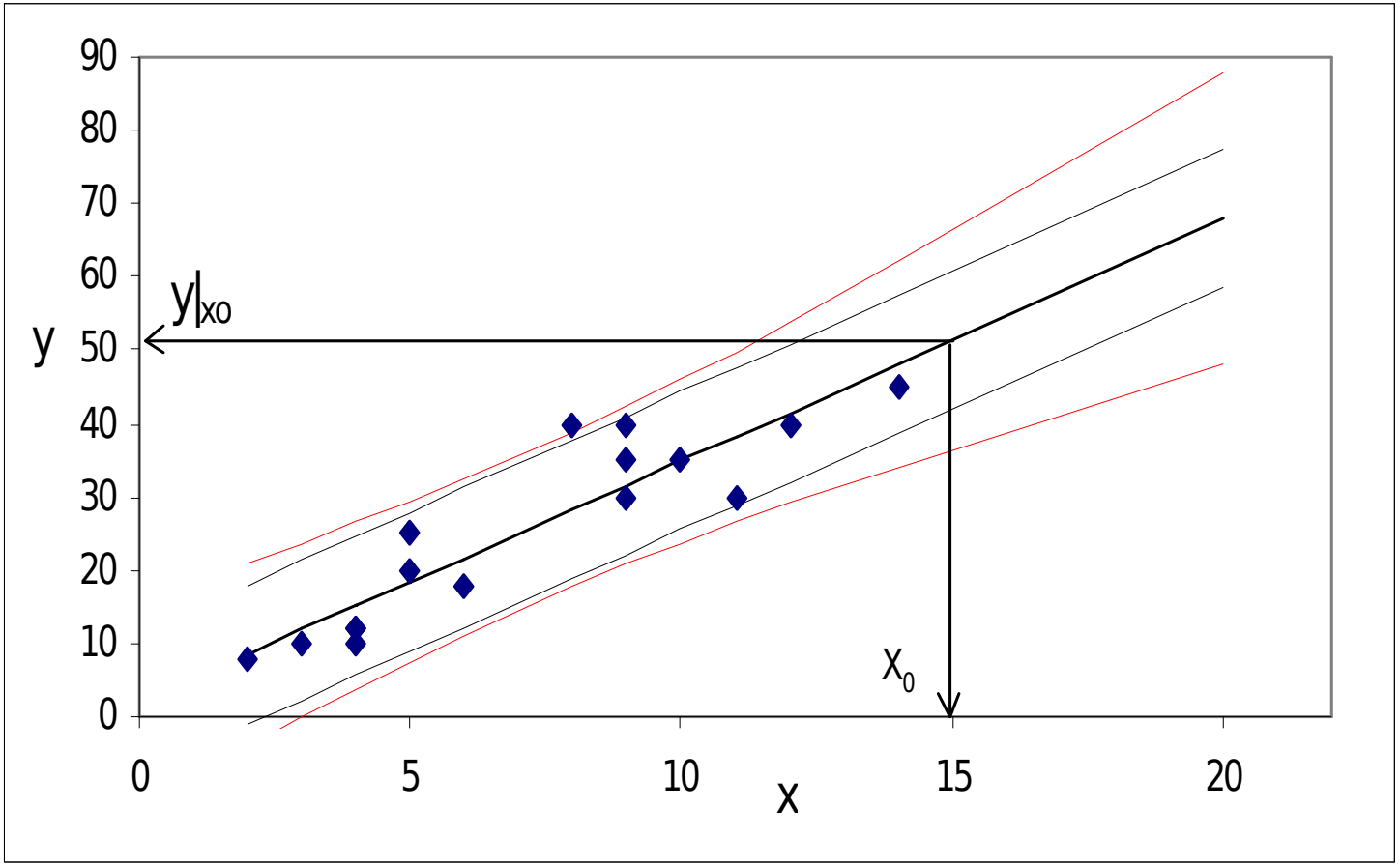
Where:

$\mathbf{x_0}$: represents a set of observations of *x's*

$\mathbf{X}$: consists of historical data of *x's*

$(\mathbf{X}'\mathbf{X})^{-1}$: Proportional to variance-covariances of regression coefficients

Confidence Interval



Calculation of s_{Pred} in Excel - an example

- Mica Watershed July 1st forecast
- *y* variable:
  - Jul-Sep residual runoff volume
- *x* variables:
  - x_1: Mt Abbot May 1st snow water equivalent (mm)
  - x_2: Kootenay NP KTNY Crosing winter (Oc-Ap) precip (mm)
  - x_3: Glacier NP Mt Fidelity mean max temp for June ^{o}C)
- Historical data:
  - 1961-94: 34 years of data